

Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity

Year: 2021 | Citations: 1343 | Authors: Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, Pontus Stenetorp

Abstract

When primed with only a handful of training samples, very large, pretrained language models such as GPT-3 have shown competitive results when compared to fully-supervised, fine-tuned, large, pretrained language models. We demonstrate that the order in which the samples are provided can make the difference between near state-of-the-art and random guess performance: essentially some permutations are “fantastic” and some not. We analyse this phenomenon in detail, establishing that: it is present across model sizes (even for the largest current models), it is not related to a specific subset of samples, and that a given good permutation for one model is not transferable to another. While one could use a development set to determine which permutations are performant, this would deviate from the true few-shot setting as it requires additional annotated data. Instead, we use the generative nature of language models to construct an artificial development set and based on entropy statistics of the candidate permutations on this set, we identify performant prompts. Our method yields a 13% relative improvement for GPT-family models across eleven different established text classification tasks.